

# OPEN OBJECT RECOGNITION FOR HUMANOID ROBOTS

*Copyright 2003 Society of Photo-Optical Instrumentation Engineers. This paper will be published in Robotics and Machine Perception and is made available as an electronic preprint with permission of SPIE. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.*

Paul Fitzpatrick  
MIT AI Laboratory  
Cambridge, MA 02139, USA  
Email: paulfitz@ai.mit.edu

Robots must be able to adapt gracefully to frequent and dramatic changes in their workspace if they are to operate successfully in human-centered environments, as opposed to controlled industrial settings. At the MIT Humanoid Robotics Group, we are developing methods that permit our robots to deduce the structure of novel activities, adopt the vocabulary appropriate for communication about the task at hand, and learn about the appearance and behavior of unfamiliar objects. This latter ability is discussed here. The humanoid robot Cog [1] uses active exploration to resolve visual ambiguity in its workspace [2]. As Cog accumulates experience, it clusters episodes of object interaction to learn the appearance and properties of novel, unfamiliar objects. This process is called open object recognition [3]. An operator can then introduce names for objects to facilitate further task-related communication.

## I. ACTIVE SEGMENTATION

Figure/ground separation is a long-standing problem in computer vision, due to the fundamental ambiguities involved in interpreting the 2D projection of a 3D world. Cog can bypass this philosophical and practical dilemma by physical experimentation (see Figure 1). Cog has a 'poking' behavior that prompts it to select locations in its environment that may contain an object of interest, and sweep through them with its arm [2]. If an object is within the area swept, then the motion generated by the impact of the arm can be used to segment the object from its background, and obtain a reasonable estimate of its boundary. This is called active segmentation, and is a form of active perception [4]. Once Cog can reliably segment objects, then it learns about their appearance and how they move. Of course, active segmentation does not work for all objects – if an object is very small or very large, the procedure is likely to fail. But manipulable objects are almost by definition on the right scale for the method to work, and this is a particularly important class of object for robots.

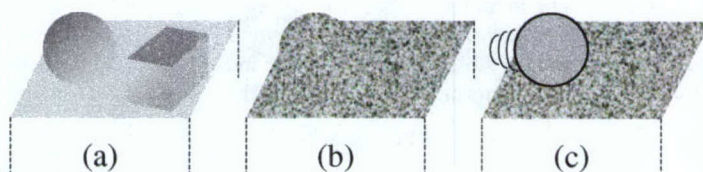


Figure 1: Cartoon motivation for active segmentation. Human vision is excellent at figure/ground separation (top left), but machine vision is not (center). Coherent motion is a powerful cue (right) and the robot can invoke it by simply reaching out and poking around.

## II. OPEN OBJECT RECOGNITION

Open object recognition is the ability to recognize a flexible set of objects, where new objects can be introduced at any time [3]. Cog can learn autonomously to recognize new objects by interacting with them (see Figure 2). Conventional object recognition systems do not need to be open - for example, the set of objects an industrial robot needs to interact with is likely to be fixed. But a humanoid robot in an unconstrained environment could be presented with just about anything, and trying to



collect and train for all the possible objects the robot might encounter is simply not practical. Active segmentation gives Cog the ability to collect its own training data for machine learning. A variant of geometric hashing is used for object localization, with clustering of object models occurring both on- and off-line. The online clustering procedure is fast and responsive (on the order of seconds), but relatively coarse. The off-line clustering procedure is slower (on the order of tens of minutes), but can make subtler distinctions between objects. Both clustering methods are integrated so that the robot can distinguish visually distinctive objects quickly and more difficult cases over time.

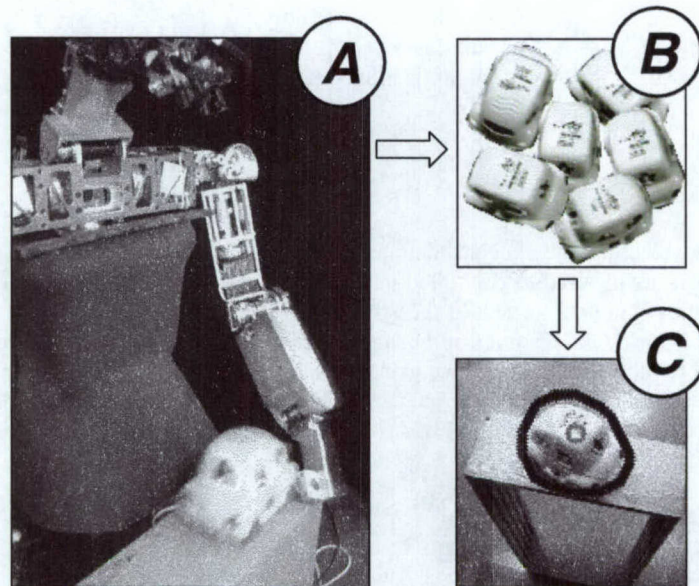


Figure 2: Object boundaries are not always easy to detect visually. The robot Cog (A) solves this by sweeping its arm through areas of ambiguity. If object motion results, the motion helps distinguish the object from its background (B). As the robot gains experience and becomes familiar with the appearance of an object, it learns to recognize and segment that object without further contact (C).

### III. CONCLUSION

The methods touched upon here allow our humanoid robot Cog to build up and maintain a perceptual system for object localization, segmentation, and recognition, starting from very little. Beyond this, Cog can track known objects to learn about activities they occur in, such as a sorting task or object search [3]. The overall goal of this effort is to develop a perceptual system for a humanoid robot that is as general-purpose and adaptable as the robot's physical form.

### REFERENCES

- [1] R. A. Brooks, C. Breazeal, M. Marjanovic, and B. Scassellati, The Cog project: building a humanoid robot, Lecture Notes in Artificial Intelligence 1562, C. Nehaniv (ed.), pp. 52-87, Springer, 1999
- [2] P. Fitzpatrick and G. Metta, Grounding vision through experimental manipulation, Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences, in press.
- [3] P. Fitzpatrick, From first contact to close encounters: a developmentally deep perceptual system for a humanoid robot. PhD thesis and technical report AITR-2003-008, MIT Department of Electrical and Computer Engineering, 2003
- [4] R. Bajcsy, Active perception, Proceedings of the IEEE 76(8), pp. 966-1005, August 1988